Yue Xin

📞 (+86) 19821286629　🎨 Machine-Phoenix　✉ xinyuexiong@sjtu.edu.cn　🔗 [Yue Xin](#)

SUMMARY

Research Interests: My research primarily lies in machine learning, interpretable AI, and large (language) models. My goal is to identify the fundamental optimization goal of deep learning and establish interpretable machine learning systems.

Experience: First author at top conferences; Internship experience at renowned internet companies; Top 5% in Grades; National Scholarship; Mathematics and Physics competitions

Skills: Solid mathematical background; over 3 years of research experience; over 5 years of programming experience.

EDUCATION

Shanghai Jiao Tong University (SJTU)　　　　　　　　　　　　　　　　　　**June 2023 – Present**
M.S. in Information and Communication Engineering　　　　　　　　　　　　　　*Shanghai, China*

- Adviser: Prof. Hongkai Xiong and Prof. Wenrui Dai

Shanghai Jiao Tong University (SJTU)　　　　　　　　　　　　　　　　　　**Sep 2019 – June 2023**
B.S. in Electronic Science and Technology (major) and Computer Science and Technology (minor)　　　　*Shanghai, China*

- Upon graduation (all courses):　　　　　　　　　Grades: 89.66/100　　GPA: 3.85/4.3　　Ranking: 4/56
  When receiving the National Scholarship (all courses):　Grades: 90.10/100　　GPA: 3.89/4.3　　Ranking: 2/64
- A(90+) for all math/physics related courses(Calculus, Physics, Probability and Statistics, etc) and most major related courses(Signals and Systems, Progamming, Digital Image Processing, etc). A+(95+) for most major related experimental courses(Engineering Problem Modeling and Simulation, Engineering Practice and Technological Innovation).

PUBLICATIONS

- **Clarifying the Behavior and the Difficulty of Adversarial Training**
  Xu Cheng*, Hao Zhang*, **Yue Xin**, Wen Shen, Jie Ren, Quanshi Zhang.
  AAAI 2024 accepted. [paper]

- **Towards the Dynamics of a DNN Learning Symbolic Interactions**
  Qihan Ren, Yang Xu, Junpeng Zhang, **Yue Xin**, Dongrui Liu, Quanshi Zhang
  NeurIPS 2024 accepted [paper]

- **ChebHiPoly: Hierarchical Chebyshev Polynomial Modules for Enhanced Approximation and Optimization**
  Yue Xin, Jiarui Zhang, Ziyang Zheng, Yaoming Wang, Wenrui Dai, Chenglin Li, Junni Zou, Hongkai Xiong.
  ICME 2025 on submission

- **SalaMAnder: Shapley-based Mathematical Expression Attribution and Metric for Chain-of-Thought Reasoning**
  Yue Xin, Chen Shen, Shaotian Yan, Yaoming Wang, Xiaosong Yuan, Chenxi Huang, Jieping Ye.
  ACL 2025 on submission

- **RST: Residual Side Tuning with Cross-Layer Correlation for Parameter Efficient Transfer Learning**
  Yue Xin, Yaoming Wang, Wenrui Dai, Jiarui Zhang, Ziyang Zheng, Chenglin Li, Junni Zou, Hongkai Xiong.
  ICML 2025 on submission

- **GLEAM: Global Share Local Transform MoE for Downstream Transferring With Enhanced Parameter Efficiency**
  Jiarui Zhang, **Yue Xin**, Yaoming Wang, Ziyang Zheng, Wenrui Dai, Chenglin Li, Junni Zou, Hongkai Xiong.
  ICME 2025 on submission

- **Bootstrap Prompt Learning with Feature Adaptation for Vision-Language Efficient Tuning**
  Jiarui Zhang, Yaoming Wang, **Yue Xin**, Wenrui Dai, Ziyang Zheng, Chenglin Li, Junni Zou, Hongkai Xiong.
  ICML 2025 on submission

- **Enhancing Chain-of-Thought Reasoning with Critical Representation Fine-tuning**
  Chenxi Huang, Liang Xie, Chen Shen, Shaotian Yan, Sinan Fan, **Yue Xin**, Binbin Lin, Deng Cai, Jieping Ye.
  ACL 2025 on submission

ACADEMIC RESEARCH EXPERIENCE

Institute of Media, Information and Network(min), SJTU   Nov 2022 – Present
Machine Learning and Computer Vision Intern and Master's Student   *Advisor: Hongkai Xiong, Wenrui Dai*
- Proposed modular Chebyshev connections into general network layers to improve the approximation capability of a neural network. Specifically, established recursive relationship among adjacent layers and polynomial relationship between non-adjacent layers. Comprehensive experiments vefify its strong approximation and optimization capability. (ICME 2025 on submission)
- Proposed Residual Side Tuning (RST) framework to enhance information extraction efficiency by employing a dual-block side-tuning structure with low-rank linear mapping on aggregated features, and introducing element-wise feature enhancement strategy to model cross-layer information. The properties and performance of RST is verified through mathematical proof and various experiments. (ICML 2025 on submission)
- Proposed GLEAM, an efficient fine-tuning method for large model parameters. This method leverages the high similarity of parameter matrices in LoRA to construct a low-rank decomposition, further reducing the number of parameters required for fine-tuning while enhancing performance. (ICME 2025 on submission)
- Proposed PAT, a fault-tolerant multimodal classifer to solve the problem of conflict between prompter learning and adapter tuning. Specifically, it aligns model representations obtained from Soft Prompt and Adapter-based methods and incorporats contrastive learning loss to enhance model performance and generalization. (ICML 2025 on submission)

Feitian Lab, Alibaba Cloud   Mar 2024 – Oct 2024
Interpretable LLM Research Intern   *Advisor: Jieping Ye*
- Proposed SalaMAnder, a Shapley-value-based framework for quantifying component-level contributions in CoT reasoning. Specifically, we develop an efficient stratified sampling algorithm to compute Shapley value for mathematical expression attribution and CoSP (Cardinality of Shapley Positives) metric. Theoretical derivation and comprehensive validation across multiple models and benchmarks present a robust monotonic correlation between CoSP and model performance, providing theoretical explanations for the empirical success of CoT. (ACL 2025 on submission)
- Classfied and summarized research work on interpretability on large language models.

Interpretable ML lab, SJTU   Feb 2022 – Nov 2022
Interpretable Machine Learning Intern   *Advisor: Quanshi Zhang*
- Theoretically derived the analytical solution for multi-step adversarial attacks, which explains the reasons behind the optimization difficulties in adversarial training. This is validated through experiments. (Accepted by AAAI 2024)
- Theoretically derived the two-stage dynamic interaction process of DNNs, proving that the network learning process gradually encodes interactions of varying complexity. This provides a theoretical foundation for understanding overfitting. (Accepted by NeurIPS 2024)
- Theoretically derived and validated the robustness of concepts with different complexities.

ACADEMIC COMPETITION (Selected)

The 20th Chinese Graduate Mathematical Modeling Competition:	*Nation level, Second Prize*	**2023**
The Mathematical Contest in Modeling:	*World level, Meritorious Winner (First Prize)*	**2021**
The Huawei Cloud 'Cloud Pioneers' Few-Shot Detection Competition:	*Nation level, Third Place*	**2021**
The 12th National College Student Mathematical Competition:	*City level, First Prize*	**2020**
The 2nd National 'August 1st Cup' Online Mathematics Competition:	*Nation level, Tenth Place*	**2020**
Chinese Physics Olympiad:	*Province level, First Prize*	**2018**

TECHNICAL SKILLS

Programming Languages: Proficient in Python, C++, Matlab, LaTeX, Linux, etc.

Frameworks: Proficient in PyTorch, NumPy, Anaconda, Git, OpenCV.

Mathematics: Proficient in calculus, linear algebra, probability statistics, etc.

Language: mandarin (native), English (fluent)

HONORS & AWARDS (Selected)

Outstanding Undergraduate Graduate of Shanghai Jiao Tong University	*University level, 5%*	**2023**
National Scholarship	*Nation level, 2%*	**2021**
Shanghai Jiao Tong University A-Class Excellent Scholarship for Undergraduate	*University level, 2%*	**2021**
Shenzhen Stock Exchange Scholarship	*University level, 2%*	**2020**
Shanghai Jiao Tong University B-Class Excellent Scholarship for Undergraduate	*University level, 5%*	**2020**